November 18, 2024

Hon. Ona T. Wang
United States Magistrate Judge
Southern District of New York

>   Re: *The New York Times Company v. Microsoft Corporation, et al.*, 23-cv-11195; *Daily News, L P, et al. v. Microsoft Corp., et al.*, 1:24-cv-3285; *The Center for Investigative Reporting, Inc. v. OpenAI, Inc.*, 1:24-cv-04872: Dispute Regarding Early Deposition on Microsoft's Log Data

Dear Magistrate Judge Wang:

I write on behalf of Plaintiffs Daily News, LP et al., The New York Times Company, and The Center for Investigative Reporting, Inc. (collectively the "News Plaintiffs"). We ask that you compel Microsoft to provide a witness for a 30(b)(6) deposition prior to December 20, 2024, on the topic of Microsoft's log data reflecting user sessions and associated output logs. This 30(b)(6) deposition should take place without prejudice to the News Plaintiffs taking a subsequent 30(b)(6) deposition at a later point in discovery on additional topics. The purpose of this early 30(b)(6) deposition is to streamline discovery on issues important to the News Plaintiffs' case, as explained further below.

On November 1, 2024, the parties submitted a joint letter at the direction of the Court about their efforts to arrange an early 30(b)(6) custodial deposition. *See generally* Dkt. 306.[1] The request in this motion does not override the News Plaintiffs' request in that letter for other deposition topics, but overlaps in substance with topics 15 and 16 in Exhibit 1 attached to that letter, which specifically deal with log issues:

> 15. The existence, organization, preservation, storage, and deletion of output data reflecting user sessions on Your Generative AI Products and Services, as well as of reports on, metrics on, or analysis of such output data.

> 16. The meaning of data fields contained in sample output logs produced by Microsoft at MSFT_NEWS_000624874, MSFT_NEWS_000624875, and MSFT_NEWS_000624876, as well as of data fields contained in other output logs maintained by Defendants.

Dkt. 306-1 at 2.

In the joint letter, Microsoft said that a deposition on the log issues is "not a 'custodial' deposition" and "will happen in due course, but there is no reason to upend the normal case schedule and rush into it now." Dkt. 306 at 3. More recently, Microsoft has refused to provide a witness by the interim fact discovery cutoff of December 20, 2024. Ex. A. But the News Plaintiffs respectfully submit that ***now*** is precisely the right time for such a deposition so that the parties may streamline discovery into issues that are important to the News Plaintiffs' case.

---

[1] All docket references herein are to case no. 23-cv-11195.

On May 22, 2024, and June 14, 2024, the News Plaintiffs first served requests for inspection pursuant to Fed. R. Civ. P. 34 on Microsoft (and OpenAI) in order to ascertain Defendants' unauthorized use of the News Plaintiffs' copyrighted content through their artificial intelligence systems. Exs. B-E. This use includes, *inter alia*, Defendants copying the News Plaintiffs' content as part of their training datasets (ingestion), encoding of the News Plaintiffs' content within the parameters of the large language models themselves through training (digestion), copying the News Plaintiffs' content to perform retrieval-augmented generation ("RAG"), and including the News Plaintiffs' content in generative output to end-users because of model "memorization" or RAG capabilities (regurgitation). Discovery into some of these technical issues is not amenable to traditional document production mechanisms because of the volume of data at issue and/or the need for specialized computing resources that are solely in the Defendants' possession.

The parties have been negotiating the News Plaintiffs' requests for months, and have made some progress navigating these technical areas of discovery—but not without many, many disputes.[2] Microsoft produced "sample" log data and documentation to the News Plaintiffs just before the October 30 status conference reflecting the fields of data that it maintains for its Bing Chat product.[3] This data, including the generative output provided to end-users, not only serves as evidence of copyright infringement, but also informs the damages and fair use analysis as to whether Defendants' products are competing as substitutes for the News Plaintiffs' copyrighted works.

Because of the opacity of the sample log data, and Microsoft's representations that the log data is extremely voluminous and burdensome to search, the News Plaintiffs have asked for an early 30(b)(6) deposition on issues pertaining to the log data before the December 20, 2024, interim fact discovery cutoff—without prejudice to a subsequent 30(b)(6) deposition. Microsoft has taken the position that it does not know what the deposition would be about. But the News Plaintiffs have told Microsoft that "the purpose of the deposition is to identify what documents (or data) Microsoft has or can create in order to show which of the News Plaintiffs' works have been used to create output from Microsoft's GAI products, what content has been outputted, how many times that has occurred, and in response to what queries—along with the specifics of how to search Microsoft's logs for that information." Ex. A at 1. This early deposition will allow the News Plaintiffs to more efficiently conduct discovery.

November 18, 2024   Respectfully submitted,

                                                        /s/ Ian Crosby
                                                        Ian B. Crosby

---

[2] The parties have outstanding disputes with respect to the News Plaintiffs' inspection of OpenAI's training data (Dkt. 305) and inspection of the models (Dkt. 250 at 4-6).

[3] OpenAI produced log data for a ChatGPT session with Browse on November 13, 2024, which the News Plaintiffs are still reviewing. News Plaintiffs, however, submit that a similar 30(b)(6) deposition directed to OpenAI may also significantly streamline discovery for the same reasons stated herein.

Sussman Godfrey L.L.P.

/s/ Steven Lieberman
Steven Lieberman
Rothwell, Figg, Ernst & Manbeck

/s/ Matt Topic
Matt Topic
Loevy & Loevy

cc: All Counsel of Record (via ECF)